DOCUMENT RESUME

ED 202 872 TM 810 291

AUTHOR	Friedman, David; And Others
TITLE	Adaptive Testing Without a Computer.
INSTITUTION	Research Applications, Inc., Rockville, Md.
SPONS AGENCY	Air Force Human Resources Lab., Brooks AFB, Texas.
REPORT NO	AFHRL-TR-80-66
PUB DATE	Mar 81
CONTRACT	F33615-79-C-0018
NOTE	27p.
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Aptitude Tests; Arithmetic; *Armed Forces; *Correlation; *Testing; Word Recognition
IDENTIFIERS	*Adaptive Testing; Armed Forces Qualification Test; Armed Services Vocational Aptitude Battery; *Paper and Pencil Tests; Time Utilization

ABSTRACT

Three prototypes of paper-and-pencil based adaptive tests were developed, refined and developed in sufficient quantities for administration to groups of 40 subjects. Two aptitude areas were employed in each prototype. These were Word Knowledge and Arithmetic Reasoning. A total of 711 Basic Airmen Recruits were administered the prototypes in both aptitude areas as well as traditional paper-and-pencil tests of both areas. Additionally, enlistment scores on the Armed Forces Qualification Test (AFQT) and the Armed Services Vocational Aptitude Battery (ASVAB) were available for each subject. It was found that the adaptive tests correlated highly with like-named paper-and-pencil tests and correlations with AFQT and the ASVAB were about the same for traditional tests and adaptive tests. The adaptive tests showed a large advantage in time of administration ranging from savings of one-third to one-half. It is anticipated that a full adaptive test battery based on the prototypes would allow for the addition of about six more aptitude areas. This could provide better measurement by enabling more data to be collected on each examinee. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

AFHRL-TR-80-66

AIR FORCE HUMAN RESOURCES LABORATORY

ED202872

ADAPTIVE TESTING WITHOUT A COMPUTER

By

David Friedman
Alma Steinberg
Research Applications, Incorporated
1776 East Jefferson Street
Rockville, Maryland 20852

Malcolm James Ree

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
the Office of Naval Research
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

March 1981

Final Report

Approved for public release; distribution unlimited.

TM 810 291

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by Research Applications, Incorporated, 1776 East Jefferson Street, Rockville, Maryland 20852, under Contract F33615-79-C-0018, Project ILIR, with the Manpower and Personnel Division, Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235. Malcolm James Ree was the Contract Monitor for the Laboratory.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

NANCY GUINN, Technical Director
Manpower and Personnel Division

RONALD W. TERRY, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-80-66	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* ADAPTIVE TESTING WITHOUT A COMPUTER		5. TYPE OF REPORT & PERIOD COVERED Final
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David Friedman Alma Steinberg Malcolm James Ree		8. CONTRACT OR GRANT NUMBER(s) F33615-79-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Applications, Incorporated 1776 East Jefferson Street Rockville, Maryland 20852		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61101F ILIR0017
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS*(If different from Controlling Office)* Manpower and Personnel Division Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT *(of this Report)* Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)* test and measurement latent trait adaptive testing psychometrics ASVAB selection		

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Three prototypes of paper-and-pencil based adaptive tests were developed, refined and developed in sufficient quantities for administration to groups of forty subjects. Two aptitude areas were employed in each prototype. These were Word Knowledge and Arithmetic Reasoning. A total of 711 Basic Airmen Recruits were administered the prototypes in both aptitude areas as well as traditional paper-and-pencil tests of both areas. Additionally, enlistment scores on the Armed Forces Qualification Test (AFQT) and scores for selector composites (AI) were available for each subject. It was found that the adaptive tests correlated highly with like-named paper-and-pencil tests and correlations with AFQT and the AIs were about the same for traditional tests and adaptive

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Item 20 Continued:

tests. The adaptive tests showed a large advantage in time of administration ranging from savings of one-third to one-half. It is anticipated that a full adaptive test battery based on the prototypes would allow for the addition of about six more aptitude areas. This could provide better measurement by enabling more data to be collected on each examinee.

PREFACE

This research was conducted under ILIR0017, Adaptive Testing Without A Computer.

The authors wich to express their appreciation to the testing detachment of the Manpower and Personnel Division at Lackland AFB, Texas, the Technical Service Division and James B. Sympson for contributions to this effort.

I. INTRODUCTION

Item reponse theory, often referred to as latent-trait theory, has provided the tools for solving the problem of tailoring a test to the individual. Traditionally, the same test is given to all individuals regardless of the ability level of the individual and the difficulty level of the test. This mismatch may result in decreased precision of measurement which may, in turn, lead to misclassification, errors of selection, poor use of scarce resources and selection of individuals who are ill-equipped to perform the tasks at hand.

The development of latent-trait theory (see Lord & Novick, 1968) has been the latest in a constant trend toward making human aptitude measurement more precise by adapting tests to examinees.

As early as the beginning of the twentieth century, Alfred Binet (see Peterson, 1926) developed adaptive tests for educational screening. The success of the group-administered tests developed during the first World War, coupled with the long administration time of the Binet tests, changed the course of test development to efforts aimed at producing the more economical paper-and-pencil group-administered non-adaptive measurements which have become the standard.

The advent of relatively inexpensive and portable computers has made feasible computer-directed adaptive testing. In the last decade, numerous studies have been undertaken in an attempt to accomplish adaptive measurement using computers (see Weiss, 1977).

Computers, however, are prone to failures at unpredictable times and are still more expensive than paper-and-pencil media. This effort, therefore, was designed to investigate the feasibility of developing sophisticated adaptive tests which do not rely on computer administration techniques. Such tests would eliminate the need for costly machines, capture the advantages of latent-trait theory, and be as portable as ordinary test booklets.

II. METHOD

The Adaptive Test

For this effort, an adaptive test was defined as a test composed of several scorable items which were administered sequentially, so that the item presented was based on the results of the preceding question, or on the results of all the preceding questions. In an adaptive testing environment, the examinee is routed from item to item so that not all examinees necessarily answer all questions nor necessarily the same number of questions (McBride, 1977).

Item Pools

Two adaptive content areas, Word Knowledge (WK) and Arithmetic Reasoning (AR), were used for the adaptive tests. Using the maximum likelihood procedure described by Wingersky and Lord (1973), the test items for these content areas had been calibrated on a sample of approximately 1,600 Air Force recruits. Each ability area was calibrated separately using the three-parameter logistic

model (Birnbaum, 1968). Items which had parameters out of range were deleted from the pool, leaving a set of items which were appropriate for the testing task.

Prototype Development

Five prototype strategies for adaptive testing were proposed, and three of these were selected for tryout on small samples of Air Force basic recruits to refine procedures and techniques. The prototypes were designed so that once the initial instructions were given, the subject would not require further assistance to complete the test.

A "routing test" followed by a "measurement test", was used in each prototype. These procedures resulted in a two-stage test protocol. Two methods of routing the subject from item to item were used. For one method, all subjects answered all items in the first stage of the test. Depending on their performance on the first stage, they were routed to one of five second-stage tests.

For the second routing method, all subjects started with the first item in the first stage of the test. Depending on whether their response was correct or incorrect, subjects were routed to a more or less difficult item. This same procedure was followed for each subsequent item in the first stage. The sequences of items answered determined the level of the test to be taken at the second stage.

Prototype I

In Prototype I (PI), each examinee used a cardboard box containing 450 7.62 x 12.70 cm (3 x 5-inch) item cards. The test items for the two subtests were printed on these cards. The tests were color-coded; and divider cards separated the parts of each subtest.

In order to prevent loss and disarrangement, the cards were held in a box by two rods threaded through holes in the cards and anchored by stoppers at each end. Although the cards were not to be removed from the box by the examinees, the design of the box was such that, when necessary, worn, outdated, or obsolete subtests or items could be easily replaced by the administrator.

For each of the subtests, the examinees were provided with a one-page, machine-scannable answer sheet and a separate one-page instruction sheet. The format of each answer sheet corresponded to the individual subtest, taking into account the number of questions and response options. The instruction sheet was specific to each subtest and was used by the examinees to determine the measurement subtest to be taken.

An administration manual was provided as part of the package of materials. A reusable visual display, to aid the administrator in the instruction of the examinees in the use of the prototype, and a pen with water-based ink for use with the visual display were provided.

Prototype II

Prototype II (PII) consisted of a set of two question booklets for each subtest. The questions for the first part of each subtest were presented in a small, spiral-bound booklet which contained tabbed 7.62 x 12.70 cm (3 x 5-inch) cards and cover pages. The questions for the second part of the subtest were printed in a booklet 21.52 x 27.94 cm (8 1/2 x 11 inches). The examinees were referred to the appropriate measurement test based on the directions provided on a separate one-page instruction sheet. Each examinee used a total of two sets of question booklets and instruction sheets for each administration.

The answer sheet for PII was scannable and had invisible numbers and marks precoded in the response areas. The examinees used special crayons to mark their answers. Use of these crayons revealed the previously hidden marks. One 27.94 x 43.18 cm (11 x 17-inch) answer page printed on both sides of the paper was used for the subtest.

A manual was provided for the administrator to explain the procedures to be followed in PII. A visual aid was provided to aid the administrator in explaining the routing directions for PII. The visual aid was constructed to illustrate how the hidden marks were to be revealed on the answer sheet to respond to each test item.

Prototype III

For this third prototype (PIII), the questions were presented in a 21.52 x 27.94 cm (8 1/2 x 11-inches) booklet. The responses were recorded by the examinees on a carbonless transfer answer-sheet set. Each examinee used two question booklets and carbonless transfer answer-sheet sets. Each answer-sheet set was specifically designed to correspond to a particular subtest.

A carbonless transfer answer-sheet set consisted of two pages. The top page was a machine-scannable answer sheet that was spot-glued to a second sheet of paper. The reverse side of the machine-scannable answer sheet was covered with a block pattern to inhibit reading of the second sheet, and was treated so that markings made on the answer sheet were transferred to the second page of the set. The second page provided the examinees with instructions that routed them to the appropriate measurement test based on their responses to the first part of the test.

An instruction manual for PIII was provided to the administrator. Two visual aids were used by the administrator to explain the routing scheme for PIII. Each visual aid corresponded to one page of the answer-sheet set. A pen with water-based ink was provided for use by the administrator with the visual aids.

Routing Test Development

The routing test for Prototypes I and II (PI and PII) directed the examinee from item to item depending on the response to the previous item. A maximum

information item-selection procedure was used for these two routing tests (Sympson, 1977). Items which maximized the item-information function (Birnbaum, 1968) at the estimated ability level, $\hat{\theta}$, were selected after each item was answered. Fourteen items were available in each of these tests. Figure 1 shows the possible paths through the items.

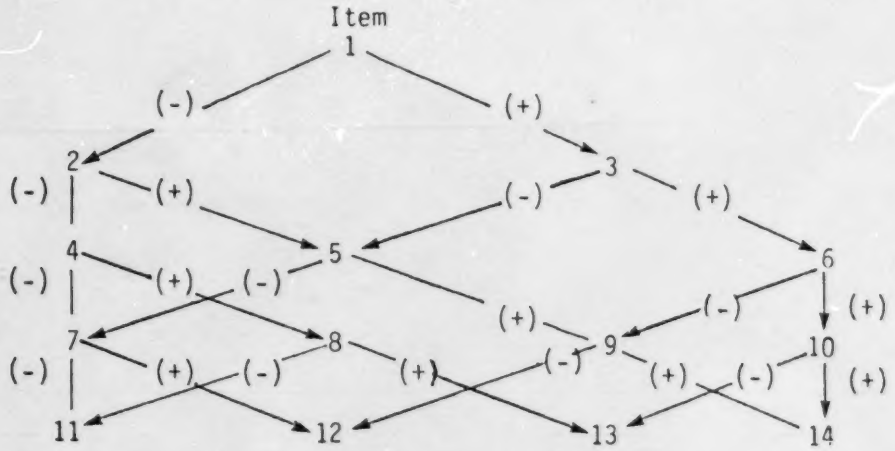

Figure 1. Paths through the routing tests for PI and PII. (Numbers indicate items; and + and - indicate correct and incorrect responses, respectively.

The routing test for Prototype III (PIII) was a short peaked measure of ability. There were eight items used in the Arithmetic Reasoning test and 10 items used in the Word Knowledge test.

Design of Administration Instructions

The administration instructions were prepared as integral parts of the prototypes. The test administrators were only to be available to reinforce these instructions or to answer appropriate questions.

The instructions were tried out with a number of volunteers whose ages ranged from nine years through adult and whose educational levels ranged from fourth grade through graduate school. On the basis of these pre-experimental trials, changes were made to the instructions in the prototypes and to the administration instructions. Instructions for the practice sessions and the special visual aids appropriate to each prototype were developed and refined. The administrators were trained in the use of these materials.

Field Test

A total of 711 airmen participated in the field test. Each took the Word Knowledge (WK) and Arithmetic Reasoning (AR) subtests from the Armed Services Vocational Aptitude Battery (ASVAB), as well as the adaptive WK and AR tests. In addition, enlistment qualification scores (scores of record) on the Mechanical, Administrative, General, and Electronics (M,A,G,E) composites of

the ASVAB, as well as the composite known as the Armed Forces Qualification Test (AFQT), were available for every subject. Other demographic data were also collected.

Instructional manuals were prepared for use by the administrators in assignment of subjects to prototype and subtest. At least 40 subjects were tested at each session. If the administrators encountered any problems at any of the sessions, they were asked to record these problems and resolutions in the manuals for review by the contractor. The initial day of administration was observed by the researchers.

For the field tryout of the prototypes, a practice test and an actual test were administered. Half of the subjects were randomly assigned to the WK adaptive tests and half were assigned the AR adaptive tests for the practice test. For the actual testing session the assignment of subjects to an adaptive test were reversed. Those subjects who were assigned the WK adaptive test for the practice session took the AR adaptive test during the actual testing session and vice versa. Thus, for each testing session, two adaptive tests were administered to each subject, one for practice and one for actual scoring.

Ability estimation in the routing test for PI and PII were determined from maximum-likelihood estimates of ability for each of the 32 possible combinations of right and wrong answers.

The routing test of PIII was designed so that all examinees took all items. These items were arranged within a short band and produced a peaked-test information function. The resultant ability estimate was used to route examinees to the appropriate measurement test.

Measurement Test Development

The measurement tests for PI and PII were the same. The medium for administration of each prototype differed. The tests were developed to provide maximum measurement precision within a relatively narrow range. This range was determined by the resultant $\hat{\theta}$ from the routing test. In order to ensure adequate coverage of the ability continuum, the measurement test information functions were carefully designed to overlap. Figure 2 represents the model.

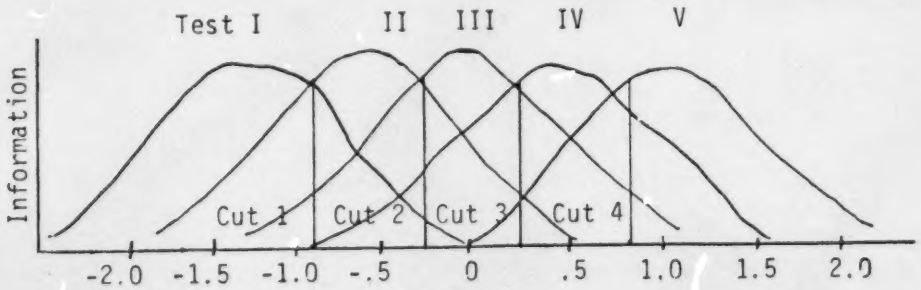

Figure 2. Overlapping information functions for measurement tests.

The measurement tests for PIII were constituted in much the same manner as for PI and PII, except that cutting points were based on the number right (NR) score. Figures 3 through 6 show the actual information functions for the measurement tests for all prototypes for both aptitude areas.

III. RESULTS

Summary statistics for age and non-adaptive WK and AR test scores were computed for the subjects. Table 1 presents these statistics for the entire group. The sample was 75 percent male and 25 percent female. Table 2 shows the average ability scores, $\hat{\theta}$, obtained by subjects for each prototype. Correlations were computed for all the variables. Tables 3, 4, and 5 show the correlations for all variables for PI, PII, and PIII.

A z test was computed (Edwards, 1958) to determine if there were differences between the correlation of the paper-and-pencil tests with AFQT and the like-named adaptive tests for AFQT. In no case were the differences significant at the predetermined $p < .05$ level.

The time required to complete the adaptive tests was recorded. ASVAB administrative times are fixed. Table 6 displays a description of the time required to complete both types of tests.

The subjects also were questioned as to their perceptions of the adaptive tests as compared to traditional paper-and-pencil tests. Table 7 presents a summary of their responses.

IV. DISCUSSION

Three prototype methods were developed to test the efficacy of the use of paper-and-pencil adaptive tests. Routing of the examinees through the test was accomplished by one of two procedures. In one routing procedure, the examinees were routed from item to item, depending on their answers to previous items. The sequence of items answered determined the second-stage level of testing. The second routing procedure provided for all the examinees to answer the same items in the first-stage test. The number of correct responses in the first stage determined the second-stage level of testing.

Two subtests (Arithmetic Reasoning and Word Knowledge) were administered to each examinee in a counterbalanced design: one for practice and one for the actual test. The items for these subtests were selected from item pools provided by the Air Force Human Resources Laboratory. ASVAB subtests in the same areas were also administered to each examinee. Examinees participated as subjects for one of three prototypes. These data were correlated with the ASVAB subtest score of the same name, and enlistment qualification composites obtained from existing records.

The results of the analyses showed that the prototype methods were successful. There was a high correlation between the ability estimates of the examinees on the subtests within each prototype and their scores on corresponding ASVAB subtests. Significance tests indicated that these observed correlations did

Figure 3: Word Knowledge Information Curves, Prototypes I and II

Cut Points 1. -1.200 3. -0.350
" θ " 2. -0.0525 4. 0.500

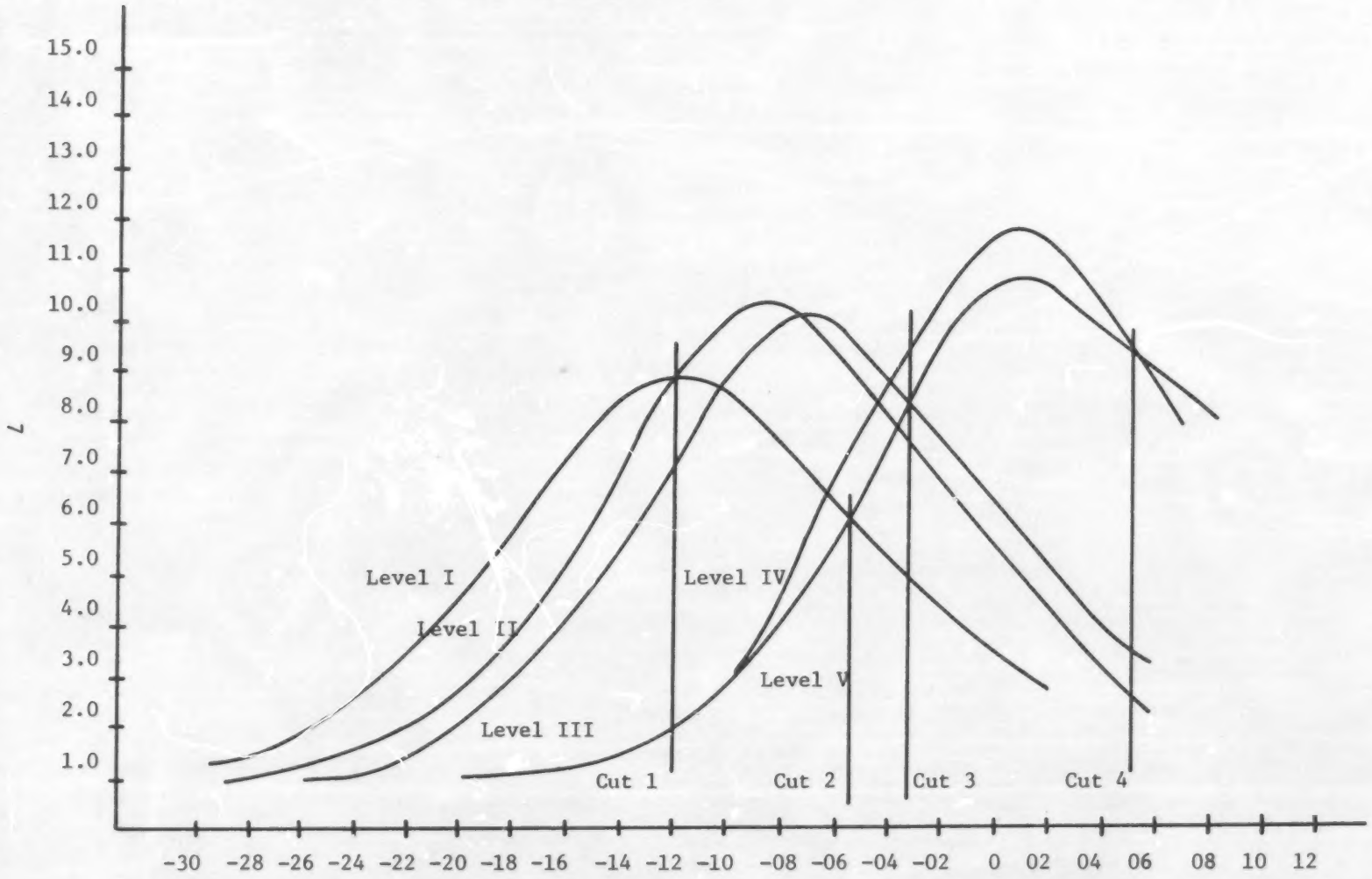

Figure 4: Arithmetic Reasoning Information Curves, Prototypes I and II

Cut Points "θ" 1. -0.85 3. 0.2
2. -0.30 4. 0.7

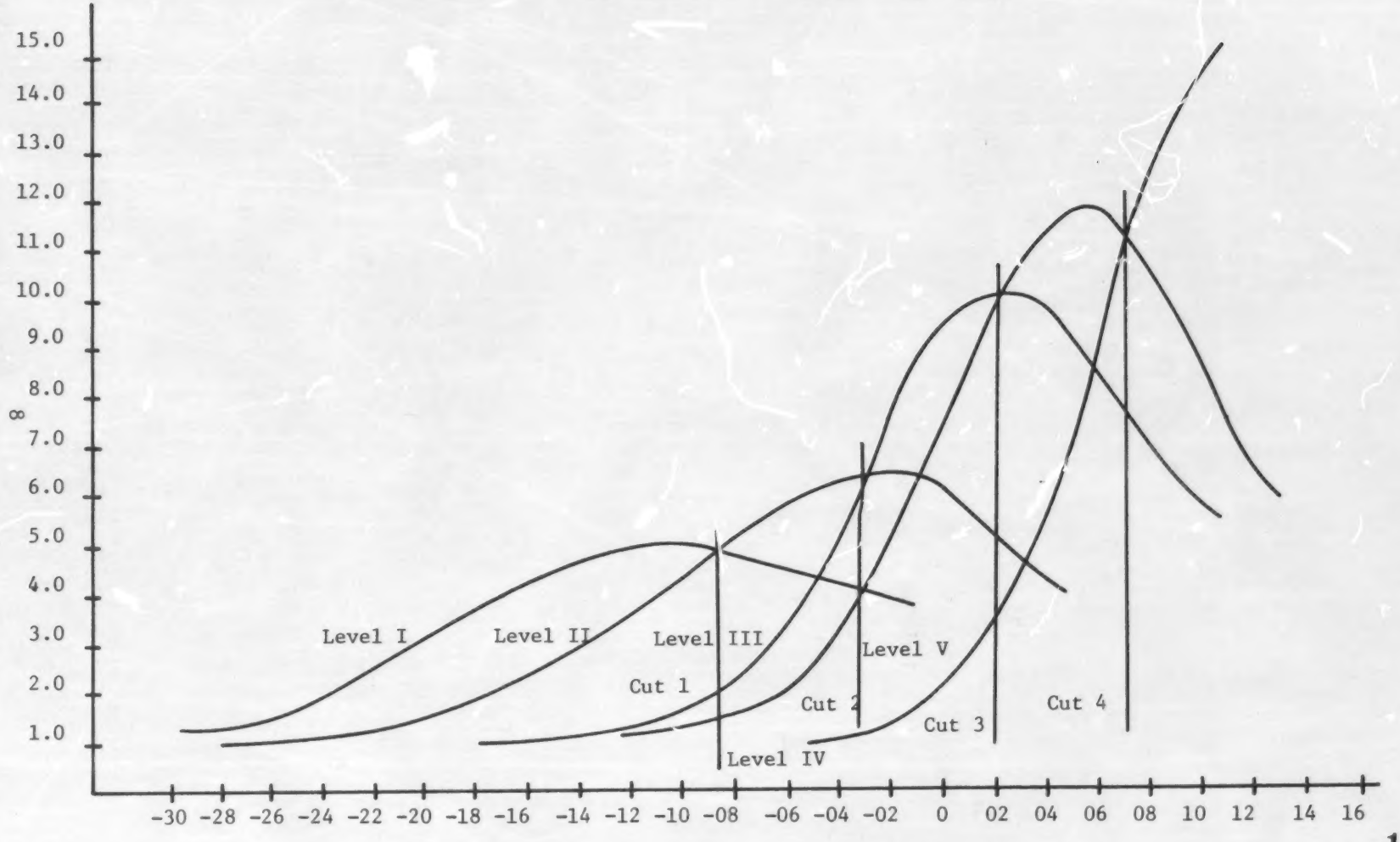

Figure 5. World Knowledge Information Curves, Prototype III

Cut Points "θ"
1. -0.8
2. -0.25
3. 0.3
4. 0.9

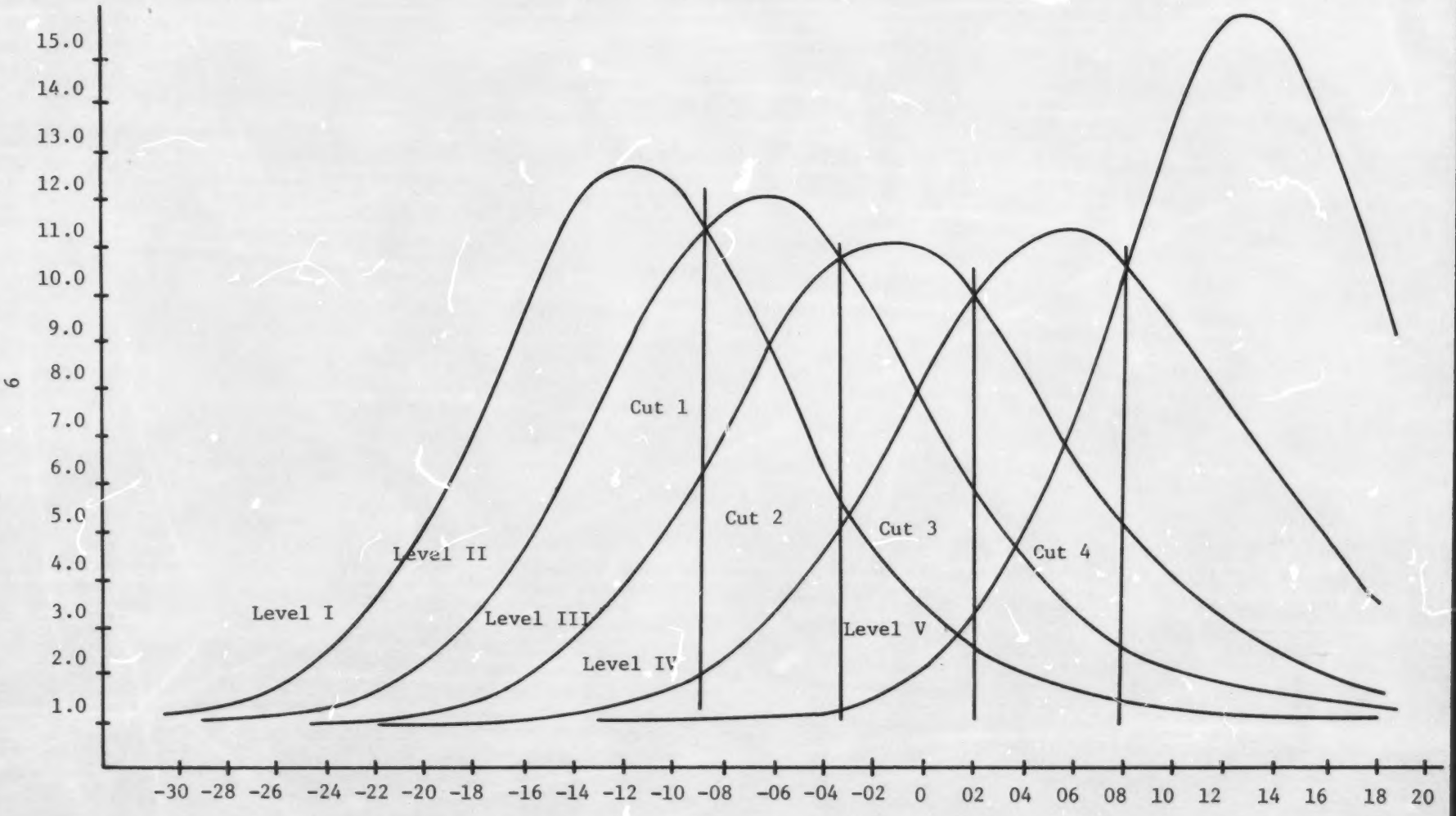

Figure 6: Arithmetic Reasoning Information Curves, Prototype III

Cut Points "θ" 1. -0.78 2. -0.20 3. 0.28 4. 0.84

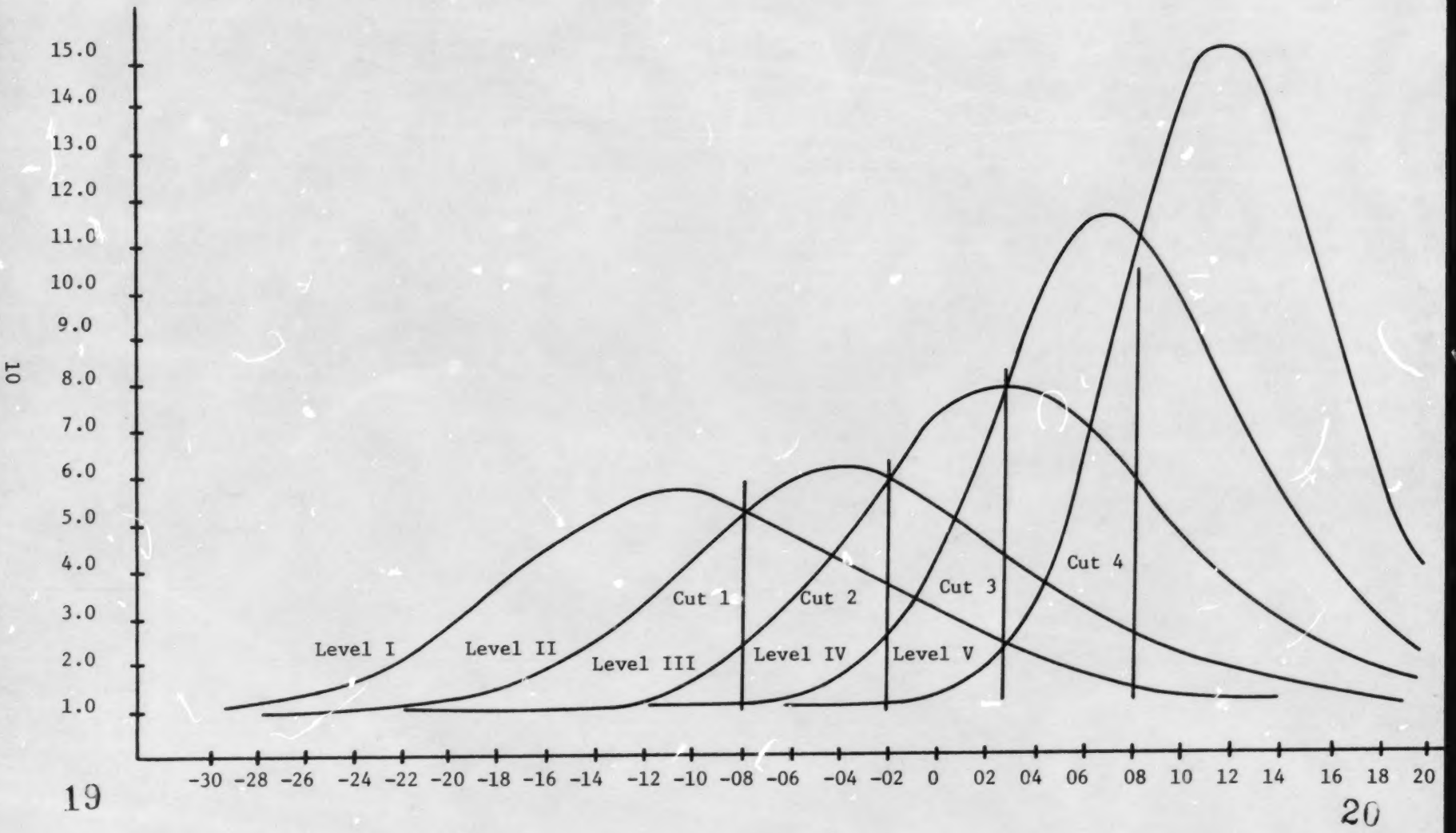

Table 1

Descriptive Statistics Age and Test Scores* for Subjects (N = 711)

Variable	Mean	Standard Deviation	Skew	Kurtosis
Age (years)	20.50	2.11	1.10	.98
AFQT	64.98	15.11	.32	- .45
M	61.29	25.05	- .05	- .96
A	69.77	19.17	- .66	- .02
G	72.56	15.16	- .30	- .80
E	71.72	17.62	- .75	- .03
ASVAB-WK	22.57	4.92	- .48	- .46
ASVAB-AR	13.90	3.91	- .03	- .67

* AFQT, M, A, G and E are reported in percentile equivalents while WK and AR are reported in number right-score.

Table 2

Descriptive Statistics for Word Knowledge and Arithmetic Reasoning Adaptive Tests.

Prototype	Aptitude	Mean	Standard Deviation	N
I	AR	-.23	.79	111
I	WK	.01	1.02	73
II	AR	-.11	.76	117
II	WK	-.02	.87	120
III	AR	-.02	.84	104
III	WK	.21	.85	67

Intercorrelations* of AFQT, Age, Sex and Test Score Variables for Prototype I.

	AFQT	AGE	SEX	$\hat{\theta}$	M	A	G	E	WK	AR
AFQT		.24	-.01	.63	.38	.40	.83	.69	.63	.63
AGE	.24		.25	.08	-.05	.24	.21	-.05	.63	.11
SEX	-.04	.14		-.09	-.62	.38	.01	-.29	-.04	-.05
$\hat{\theta}$	.77	.36	.05		.38	.42	.60	.53	.42	.74
M	.46	.15	-.52	.32		-.12	.31	.58	.36	.33
A	.44	.02	.11	.29	.06		.54	.13	.35	.40
G	.88	.20	-.04	.78	.39	.52		.50	.75	.59
E	.77	.01	-.29	.41	.61	.26	.50		.40	.49
WK	.69	.30	.03	.84	.43	.16	.71	.28		.40
AR	.63	.21	-.19	.44	.44	.41	.49	.47	.40	

*Entries above diagonal are for Arithmetic Reasoning adaptive test, $\hat{\theta}$, and those below are for the Word Knowledge adaptive test.

Table 4

Intercorrelations* of AFQT, Age, Sex, and Test Score Variables for Prototype II.

	AFQT	AGE	SEX	$\hat{\theta}$	M	A	G	E	WK	AR
AFQT		.09	-.15	.56	.39	.34	.83	.73	.66	.59
AGE	.13		.23	.06	-.09	.18	.08	-.03	.24	.01
SEX	-.06	.25		-.20	-.69	.24	-.11	-.44	-.20	-.34
$\hat{\theta}$	.72	.25	.06		.41	.31	.51	.48	.33	.68
M	.39	.00	-.66	.23		-.09	.35	.59	.37	.44
A	.35	-.01	.28	.37	-.15		.35	.10	.11	.22
G	.87	.12	-.04	.79	.30	.42		.54	.73	.61
E	.74	.01	-.44	.34	.62	.15	.59		.46	.51
WK	.64	.20	.03	.87	.26	.39	.76	.35		.43
AR	.59	.00	-.15	.51	.33	.49	.67	.62	.55	

*Entries above diagonal for Arithmetic Reasoning adaptive test, $\hat{\theta}$, and those below are for the Word Knowledge adaptive test.

Table 5

Intercorrelations* of AFQT, Age, Sex, and Test Score Variables for Prototype III.

	AFQT	AGE	SEX	$\hat{\theta}$	M	A	G	E	WK	AR
AFQT		.15	NFS**	.51	.51	.55	.84	.75	.68	.51
AGE	-.03		X	.18	.05	.21	.14	.07	.15	.14
SEX**	X	X		X	X	X	X	X	X	X
$\hat{\theta}$	.69	.06	X		.27	.44	.46	.43	.25	.73
M	.50	-.10	X	.40		.05	.35	.63	.39	.32
A	.38	.24	X	.36	.11		.63	.30	.32	.50
G	.89	.02	X	.73	.50	.35		.57	.77	.53
E	.87	-.10	X	.54	.70	.32	.81		.42	.42
WK	.70	.06	X	.85	.41	.40	.72	.59		.32
AR	.74	.02	X	.54	.43	.51	.76	.74	.59	

*Entries above diagonal are for Arithmetic Reasoning adaptive test, $\hat{\theta}$, and those below are for the Word Knowledge adaptive test

**No female subjects.

Table 6

Mean and Standard Deviation of Test Administration Times.

Test	Mean Time	Standard Deviation
ASVAB		
AR	30	*
WK	20	*
PI		
AR	21.17	5.42
WK	10.38	2.98
PII		
AR	19.67	5.10
WK	7.79	2.07
PIII		
AR	19.47	5.66
WK	8.73	2.17

*ASVAB tests of AR and WK are fixed time.

Table 7

Responses to Adaptive Versus Linear Tests

Question	Prototype I (n=232)	Prototype II (n=227)	Prototype III (n=159)
A. In your opinion, how clear were the instructions?			
1. Very clear-had no trouble at all	53.0	76.7	84.9
2. Clear enough, but could be improved	42.1	21.1	14.5
3. Unclear and confusing	4.3	2.2	0.6
4. No Response	0.4	0.0	0.0
B. How easy was it to determine the appropriate second-stage test?			
1. No trouble at all	74.1	91.6	93.1
2. A little trouble	24.6	8.4	6.3
3. A great deal of trouble	1.3	0.0	0.0
C. Did you have any problems because you wanted to make corrections, but you could not?			
1. No problems	69.4	76.2	59.1
2. Some problems	30.2	22.9	40.3
3. Serious problems	0.0	0.9	0.6
4. No response	0.4	0.0	0.0
D. How difficult were the questions?			
1. Too difficult	12.1	9.3	7.6
2. About right	84.1	86.3	87.4
3. Too easy	3.0	4.0	2.5
4. No response	0.8	0.4	2.5
E. Evaluation of this procedure for giving tests?			
1. Very good	58.6	64.8	69.8
2. Only fair	34.9	29.0	27.0
3. Poor	6.1	4.9	3.2
4. No response	0.4	1.3	0.0
F. Preference for prototype method or paper-and-pencil method of testing?			
1. Prefer prototype method	57.3	57.3	63.5
2. No difference	28.0	29.5	25.8
3. Prefer paper-and-pencil method	14.2	13.2	10.1
4. No response	0.5	0.0	0.6
G. Instructions to determine second-stage test were			
1. Very clear	56.9	78.0	76.1
2. Clear enough	39.2	20.2	22.6
3. Unclear	3.5	1.8	1.3
4. No response	0.4	0.0	0.0
H. Compared with the usual paper-and-pencil test, this method...			
1. required:			
a. more time	17.7	21.2	17.0
b. same time	21.1	23.8	22.6
c. less time	48.3	47.1	54.7
d. no response	12.9	7.9	5.7
2. was:			
a. more clear and simple	23.3	28.2	36.5
b. same in clarity and simplicity	36.6	42.7	35.8
c. less clear and simple	27.6	19.4	20.8
d. no response	12.5	9.7	6.9
3. required:			
a. more effort	26.7	18.1	19.5
b. same effort	37.6	44.1	37.1
c. less effort	22.8	30.1	37.1
d. no response	12.9	7.9	5.7
4. resulted in:			
a. more fatigue	9.0	5.7	3.8
b. same fatigue	23.7	30.8	30.8
c. less fatigue	53.9	53.4	58.5
d. no response	13.4	10.1	6.9
5. had:			
a. more questions that were matched to examinee's ability	37.5	37.0	45.9
b. same number of questions that were matched to examinee's ability	41.0	42.3	35.2
c. fewer questions that were matched to examinee's ability	9.1	11.9	13.2
d. no response	12.4	8.8	5.7
6. was:			
a. more fair	37.5	38.3	51.6
b. the same	39.7	46.3	30.8
c. less fair	9.9	7.1	11.3
d. no response	12.9	8.3	6.3
7. was:			
a. more accurate	44.0	37.9	47.2
b. the same	33.2	41.9	38.4
c. less accurate	9.5	11.0	8.8
d. no response	13.3	9.2	5.6
8. contained:			
a. more questions	2.6	4.4	1.9
b. the same number of questions	16.8	24.2	20.1
c. fewer questions	67.2	63.0	70.4
d. no response	13.4	8.4	7.6
9. offered:			
a. more opportunity to go back and review answers	28.5	31.7	28.3
b. same opportunity	25.9	34.4	23.3
c. less opportunity	33.2	26.4	42.8
d. no response	12.4	7.5	5.6
10. had:			
a. more problems in following directions	32.3	18.1	21.4
b. same problems	28.5	37.4	28.3
c. fewer problems	26.7	36.6	44.0
d. no response	12.5	7.9	6.3

not differ. The adaptive tests and the linear tests appear to be measuring the same aptitude.

Savings were obtained in the average time required to complete the adaptive tests as compared to the conventional paper-and-pencil test. The Arithmetic Reasoning (AR) subtest and the Word Knowledge (WK) subtest represent the item types which usually require the most and least time per item to administer, respectively. Reduction in AR time was about 66 percent of the usual required time, while WK time was reduced to less than half the usual time. A fully adaptive battery could be expected to allow for an increase of six subtests given in the same time required to administer Forms 6 and 7 of the ASVAB. This would provide superior measurement by enabling more data to be collected on each examinee. Reduction in classification decision errors would devolve from this additional information.

Examinees responses to the questions on perceptions about the use of adaptive testing prototypes were generally favorable, as has been found elsewhere (Prestwood & Weiss, 1978). These methods allowed them to be tested at their own level of ability and to proceed at their own rate. In addition, many felt that this kind of testing was easier than traditional testing because there were fewer items to answer, and the test taking was less fatiguing than traditional methods.

This effort provides a successful demonstration that adaptive testing can be conducted without the use of expensive computers. Further exploration and development with other aptitude areas and with a traditional criterion will have to be accomplished before any long-range decisions are made about the general implementation of these methods in the Armed Forces testing program.

REFERENCES

Birnbaum, A. Statistical theories of mental test scores. Lord and Novick (Editors) Reading, Massachusetts: Addison-Wesley, 1968.

Edwards, A.L. Statistical methods for the behavioral sciences. New York: Rinehart and Company, 1958.

Lord, F.M. & Novick, M. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

McBride, J. A brief overview of adaptive testing. (Research Report 77-1) Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977.

Peterson, J. Early Conceptions: tests of intelligence. New York: World Book Company, 1926.

Prestwood, J.S. & Weiss, D.J. The effects of knowledge of results and test difficulty on ability test performance and psychological reactions to testing. (Research Report 78-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1978.

Sympson, J.B. Estimation of latent-trait status in adaptive testing procedures. (Research Report 77-1) Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977.

Weiss, D.J. Applications of computerized adaptive testing. (Research Report 77-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, March 1977.

Wingersky, M., & Lord, F. A computer program for estimating examinee ability and item characteristic curve parameters when there are omitted responses. (Research Memorandum 73-2). Princeton, N.J.: Educational Testing Service, 1973.